1. Probability Topics: Introduction
2. Probability Topics: Terminology
3. Probability Topics: Independent & Mutually Exclusive Events
4. Probability Topics: Contingency Tables
5. Probability Topics: Summary of Formulas
6. Probability Topics: Practice with Contingency Tables
7. Probability Topics: Homework
8. Probability Topics: Review

Probability Topics: Introduction
This module introduces the concept of Probability, the chance of an event occurring.

Student Learning Objectives

By the end of this chapter, the student should be able to:

- Understand and use the terminology of probability.
- Determine whether two events are mutually exclusive and whether two events are independent.
- Construct and interpret Contingency Tables.

Introduction

It is often necessary to "guess" about the outcome of an event in order to make a decision. Politicians study polls to guess their likelihood of winning an election. Teachers choose a particular course of study based on what they think students can comprehend. Doctors choose the treatments needed for various diseases based on their assessment of likely results. You may have visited a casino where people play games chosen because of the belief that the likelihood of winning is good. You may have chosen your course of study based on the probable availability of jobs.

You have, more than likely, used probability. In fact, you probably have an intuitive sense of probability. Probability deals with the chance of an event occurring. Whenever you weigh the odds of whether or not to do your homework or to study for an exam, you are using probability. In this chapter, you will learn to solve probability problems using a systematic approach.

Optional Collaborative Classroom Exercise

Your instructor will survey your class. Count the number of students in the class today.

- Raise your hand if you have any change in your pocket or purse. Record the number of raised hands.
- Raise your hand if you rode a bus within the past month. Record the number of raised hands.
- Raise your hand if you answered "yes" to BOTH of the first two questions. Record the number of raised hands.

Use the class data as estimates of the following probabilities.
means the probability that a randomly chosen person in your class has change in his/her pocket or purse. means the probability that a randomly chosen person in your class rode a bus within the last month and so on. Discuss your answers.

- Find .
- Find .
- Find Find the probability that a randomly chosen student in your class has change in his/her pocket or purse and rode a bus within the last month.
- Find Find the probability that a randomly chosen student has change given that he/she rode a bus within the last month. Count all the students that rode a bus. From the group of students who rode a bus, count those who have change. The probability is equal to those who have change and rode a bus divided by those who rode a bus.

Probability Topics: Terminology
Probability: Terminology is part of the collection col10555 written by Barbara Illowsky and Susan Dean defines key terms related to Probability and has contributions from Roberta Bloom.

Probability is a measure that is associated with how certain we are of outcomes of a particular experiment or activity. An **experiment** is a planned operation carried out under controlled conditions. If the result is not predetermined, then the experiment is said to be a **chance** experiment. Flipping one fair coin twice is an example of an experiment.

The result of an experiment is called an **outcome**. A **sample space** is a set of all possible outcomes. Three ways to represent a sample space are to list the possible outcomes, to create a tree diagram, or to create a Venn diagram. The uppercase letter S is used to denote the sample space. For example, if you flip one fair coin, $\mathrm{S} = \{\mathrm{H}, \mathrm{T}\}$ where H = heads and T = tails are the outcomes.

An **event** is any combination of outcomes. Upper case letters like A and B represent events. For example, if the experiment is to flip one fair coin, event A might be getting at most one head. The probability of an event A is written $\mathrm{P}(\mathrm{A})$.

The **probability** of any outcome is the **long-term relative frequency** of that outcome. **Probabilities are between 0 and 1, inclusive** (includes 0 and 1 and all numbers between these values). $\mathrm{P}(\mathrm{A}) = 0$ means the event A can never happen. $\mathrm{P}(\mathrm{A}) = 1$ means the event A always happens. $\mathrm{P}(\mathrm{A}) = 0.5$ means the event A is equally likely to occur or not to occur. For example, if you flip one fair coin repeatedly (from 20 to 2,000 to 20,000 times) the relative fequency of heads approaches 0.5 (the probability of heads).

Equally likely means that each outcome of an experiment occurs with equal probability. For example, if you toss a **fair**, six-sided die, each face (1, 2, 3, 4, 5, or 6) is as likely to occur as any other face. If you toss a fair coin, a Head(H) and a Tail(T) are equally likely to occur. If you randomly guess the answer to a true/false question on an exam, you are equally likely to select a correct answer or an incorrect answer.

To calculate the probability of an event A when all outcomes in the sample space are equally likely, count the number of outcomes for event A and divide by the total number of outcomes in the sample space. For example, if you toss a fair dime and a fair nickel, the sample space is $\{\mathrm{HH},\ \mathrm{TH},\ \mathrm{HT},\ \mathrm{TT}\}$ where T = tails and H = heads. The sample space has four outcomes. A = getting one head. There are two outcomes $\{\mathrm{HT},\ \mathrm{TH}\}$. $\mathrm{P(A)} = \frac{2}{4}$.

Suppose you roll one fair six-sided die, with the numbers {1,2,3,4,5,6} on its faces. Let event E = rolling a number that is at least 5. There are two outcomes $\{5,\ 6\}$. $\mathrm{P(E)} = \frac{2}{6}$. If you were to roll the die only a few times, you would not be surprised if your observed results did not match the probability. If you were to roll the die a very large number of times, you would expect that, overall, 2/6 of the rolls would result in an outcome of "at least 5". You would not expect exactly 2/6. The long-term relative frequency of obtaining this result would approach the theoretical probability of 2/6 as the number of repetitions grows larger and larger.

This important characteristic of probability experiments is the known as the **Law of Large Numbers**: as the number of repetitions of an experiment is increased, the relative frequency obtained in the experiment tends to become closer and closer to the theoretical probability. Even though the outcomes don't happen according to any set pattern or order, overall, the long-term observed relative frequency will approach the theoretical probability. (The word **empirical** is often used instead of the word observed.) The Law of Large Numbers will be discussed again in Chapter 7.

It is important to realize that in many situations, the outcomes are not equally likely. A coin or die may be **unfair**, or **biased** . Two math professors in Europe had their statistics students test the Belgian 1 Euro coin and discovered that in 250 trials, a head was obtained 56% of the time and a tail was obtained 44% of the time. The data seem to show that the coin is not a fair coin; more repetitions would be helpful to draw a more accurate conclusion about such bias. Some dice may be biased. Look at the dice in a game you have at home; the spots on each face are usually small holes carved out and then painted to make the spots visible. Your dice may or may not be biased; it is possible that the outcomes may be affected by the slight weight differences due to the different numbers of holes in the faces. Gambling casinos have a lot of money depending on outcomes from rolling dice, so casino dice are made differently to eliminate bias. Casino dice have flat faces; the holes are completely filled with paint having the same density as the material that the dice are made out of so that each face is equally likely to occur. Later in this chapter we will learn techniques to use to work with probabilities for events that are not equally likely.

"OR" Event:
An outcome is in the event $A\text{ OR }B$ if the outcome is in A or is in B or is in both A and B. For example, let $\mathrm{A} = \{1, 2, 3, 4, 5\}$ and $\mathrm{B} = \{4, 5, 6, 7, 8\}$. $A\text{ OR }\mathrm{B} = \{1, 2, 3, 4, 5, 6, 7, 8\}$. Notice that 4 and 5 are NOT listed twice.

"AND" Event:
An outcome is in the event $\mathrm{A\ AND\ B}$ if the outcome is in both A and B at the same time. For example, let A and B be $\{1, 2, 3, 4, 5\}$ and $\{4, 5, 6, 7, 8\}$, respectively. Then $\mathrm{A\ AND\ B} = \{4, 5\}$.

The **complement** of event A is denoted $\mathrm{A'}$ (read "A prime"). $\mathrm{A'}$ consists of all outcomes that are **NOT** in A. Notice that $\mathrm{P(A)} + \mathrm{P(A')} = 1$. For example, let $\mathrm{S} = \{1, 2, 3, 4, 5, 6\}$ and let $\mathrm{A} = \{1, 2, 3, 4\}$. Then, $\mathrm{A'} = \{5, 6\}$. $\mathrm{P(A)} = \frac{4}{6}$, $\mathrm{P(A')} = \frac{2}{6}$, and $\mathrm{P(A)} + \mathrm{P(A')} = \frac{4}{6} + \frac{2}{6} = 1$

The **conditional probability** of A given B is written $\mathrm{P(A|B)}$. $\mathrm{P(A|B)}$ is the probability that event A will occur given that the event B has already occurred. **A conditional reduces the sample space**. We calculate the probability of A from the reduced sample space B. The formula to calculate $\mathrm{P(A|B)}$ is

$$\mathrm{P(A|B)} = \frac{\mathrm{P(A\ AND\ B)}}{\mathrm{P(B)}}$$

where $\mathrm{P(B)}$ is greater than 0.

For example, suppose we toss one fair, six-sided die. The sample space $\mathrm{S} = \{1, 2, 3, 4, 5, 6\}$. Let A = face is 2 or 3 and B = face is even (2, 4, 6). To calculate $\mathrm{P(A|B)}$, we count the number of outcomes 2 or 3 in the sample space $\mathrm{B} = \{2,\,4,\,6\}$. Then we divide that by the number of outcomes in B (and not S).

We get the same result by using the formula. Remember that S has 6 outcomes.

$$\mathrm{P(A|B)} = \frac{\mathrm{P(A\ and\ B)}}{\mathrm{P(B)}} = \frac{\text{(the number of outcomes that are 2 or 3 and even in S) / 6}}{\text{(the number of outcomes that are even in S) / 6}} = \frac{1/6}{3/6} = \frac{1}{3}$$

Understanding Terminology and Symbols

It is important to read each problem carefully to think about and understand what the events are. Understanding the wording is the first very important step in solving probability problems. Reread the problem several times if necessary. Clearly identify the event of interest. Determine whether there is a condition stated in the wording that would indicate that the probability is conditional; carefully identify the condition, if any.

Exercise:

Problem:

In a particular college class, there are male and female students. Some students have long hair and some students have short hair. Write the **symbols** for the probabilities of the events for parts (a) through (j) below. (Note that you can't find numerical answers here. You were not given enough information to find any probability values yet; concentrate on understanding the symbols.)

- Let F be the event that a student is female.
- Let M be the event that a student is male.
- Let S be the event that a student has short hair.
- Let L be the event that a student has long hair.

- **a** The probability that a student does not have long hair.
- **b** The probability that a student is male or has short hair.
- **c** The probability that a student is a female and has long hair.
- **d** The probability that a student is male, given that the student has long hair.

- **e** The probability that a student has long hair, given that the student is male.
- **f** Of all the female students, the probability that a student has short hair.
- **g** Of all students with long hair, the probability that a student is female.
- **h** The probability that a student is female or has long hair.
- **i** The probability that a randomly selected student is a male student with short hair.
- **j** The probability that a student is female.

Solution:

- **a** P(L')=P(S)
- **b** P(M or S)
- **c** P(F and L)
- **d** P(M|L)
- **e** P(L|M)
- **f** P(S|F)
- **g** P(F|L)
- **h** P(F or L)
- **i** P(M and S)
- **j** P(F)

**With contributions from Roberta Bloom

Glossary

Conditional Probability	The likelihood that an event will occur given that another event has already occurred.
Equally Likely	Each outcome of an experiment has the same probability.

Experiment

A planned activity carried out under controlled conditions.

Event

A subset in the set of all outcomes of an experiment. The set of all outcomes of an experiment is called a **sample space** and denoted usually by S. An event is any arbitrary subset in **S**. It can contain one outcome, two outcomes, no outcomes (empty subset), the entire sample space, etc. Standard notations for events are capital letters such as A, B, C, etc.

Outcome (observation)

A particular result of an experiment.

Probability

A number between 0 and 1, inclusive, that gives the likelihood that a specific event will occur. The foundation of statistics is given by the following 3 axioms (by A. N. Kolmogorov, 1930’s): Let S denote the sample space and A and B are two events in S . Then:

- $0 \leq P(A) \leq 1;$.
- If A and B are any two mutually exclusive events, then $P(A \text{ or } B) = P(A) + P(B)$.
- $P(S) = 1$.

Sample Space

The set of all possible outcomes of an experiment.

Probability Topics: Independent & Mutually Exclusive Events
Probability: Independent and Mutually Exclusive Events is part of the collection col10555 written by Barbara Illowsky and Susan Dean and explains the concept of independent events, where the probability of event A does not have any effect on the probability of event B, and mutually exclusive events, where events A and B cannot occur at the same time. The module has contributions from Roberta Bloom.

Independent and mutually exclusive do **not** mean the same thing.

Independent Events

Two events are independent if the following are true:

- $\mathrm{P(A|B) = P(A)}$
- $\mathrm{P(B|A) = P(B)}$
- $\mathrm{P(A\ AND\ B) = P(A) \cdot P(B)}$

Two events A and B are **independent** if the knowledge that one occurred does not affect the chance the other occurs. For example, the outcomes of two roles of a fair die are independent events. The outcome of the first roll does not change the probability for the outcome of the second roll. To show two events are independent, you must show **only one** of the above conditions. If two events are NOT independent, then we say that they are **dependent**.

Sampling may be done **with replacement** or **without replacement**.

- **With replacement**: If each member of a population is replaced after it is picked, then that member has the possibility of being chosen more than once. When sampling is done with replacement, then events are considered to be independent, meaning the result of the first pick will not change the probabilities for the second pick.
- **Without replacement:**: When sampling is done without replacement, then each member of a population may be chosen only once. In this case, the probabilities for the second pick are affected by the result of

the first pick. The events are considered to be dependent or not independent.

If it is not known whether A and B are independent or dependent, **assume they are dependent until you can show otherwise**.

Mutually Exclusive Events

A and B are **mutually exclusive** events if they cannot occur at the same time. This means that A and B do not share any outcomes and $\mathrm{P(A\ AND\ B)} = 0$.

For example, suppose the sample space $\mathrm{S} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Let $\mathrm{A} = \{1, 2, 3, 4, 5\}$, $\mathrm{B} = \{4, 5, 6, 7, 8\}$, and $\mathrm{C} = \{7, 9\}$. $\mathrm{A\ AND\ B} = \{4, 5\}$. $\mathrm{P(A\ AND\ B)} = \frac{2}{10}$ and is not equal to zero. Therefore, A and B are not mutually exclusive. A and C do not have any numbers in common so $\mathrm{P(A\ AND\ C)} = 0$. Therefore, A and C are mutually exclusive.

If it is not known whether A and B are mutually exclusive, **assume they are not until you can show otherwise**.

The following examples illustrate these definitions and terms.

Example:
Flip two fair coins. (This is an experiment.)
The sample space is $\{\mathrm{HH, HT, TH, TT}\}$ where T = tails and H = heads. The outcomes are HH, HT, TH, and TT. The outcomes HT and TH are different. The HT means that the first coin showed heads and the second coin showed tails. The TH means that the first coin showed tails and the second coin showed heads.

- Let A = the event of getting **at most one tail**. (At most one tail means 0 or 1 tail.) Then A can be written as $\{\mathrm{HH, HT, TH}\}$. The outcome HH shows 0 tails. HT and TH each show 1 tail.

- Let B = the event of getting all tails. B can be written as $\{\text{TT}\}$. B is the **complement** of A. So, $\text{B} = \text{A'}$. Also, $\text{P(A)} + \text{P(B)} = \text{P(A)} + \text{P(A')} = 1$.
- The probabilities for A and for B are $\text{P(A)} = \frac{3}{4}$ and $\text{P(B)} = \frac{1}{4}$.
- Let C = the event of getting all heads. $\text{C} = \{\text{HH}\}$. Since $\text{B} = \{\text{TT}\}$, $\text{P(B AND C)} = 0$. B and C are mutually exclusive. (B and C have no members in common because you cannot have all tails and all heads at the same time.)
- Let D = event of getting **more than one** tail. $D = \{\text{TT}\}$. $\text{P(D)} = \frac{1}{4}$.
- Let E = event of getting a head on the first roll. (This implies you can get either a head or tail on the second roll.) $E = \{\text{HT}, \text{HH}\}$. $\text{P(E)} = \frac{2}{4}$.
- Find the probability of getting **at least one** (1 or 2) tail in two flips. Let F = event of getting at least one tail in two flips. $F = \{\text{HT}, \text{TH}, \text{TT}\}$. $\text{P(F)} = \frac{3}{4}$

Example:

Roll one fair 6-sided die. The sample space is $\{1, 2, 3, 4, 5, 6\}$. Let event A = a face is odd. Then $\text{A} = \{1, 3, 5\}$. Let event B = a face is even. Then $\text{B} = \{2, 4, 6\}$.

- Find the complement of A, A'. The complement of A, A', is B because A and B together make up the sample space. $\text{P(A)} + \text{P(B)} = \text{P(A)} + \text{P(A')} = 1$. Also, $\text{P(A)} = \frac{3}{6}$ and $\text{P(B)} = \frac{3}{6}$
- Let event C = odd faces larger than 2. Then $C = \{3, 5\}$. Let event D = all even faces smaller than 5. Then $D = \{2, 4\}$. $\text{P(C and D)} = 0$ because you cannot have an odd and even face at the same time. Therefore, C and D are mutually exclusive events.
- Let event E = all faces less than 5. $E = \{1, 2, 3, 4\}$.
  Exercise:

Problem:

Are C and E mutually exclusive events? (Answer yes or no.) Why or why not?

Solution:

No. C = $\{3, 5\}$ and E = $\{1, 2, 3, 4\}$. $\mathrm{P(C\ AND\ E)} = \frac{1}{6}$. To be mutually exclusive, $\mathrm{P(C\ AND\ E)}$ must be 0.

- Find $\mathrm{P(C|A)}$. This is a conditional. Recall that the event C is $\{3, 5\}$ and event A is $\{1, 3, 5\}$. To find $\mathrm{P(C|A)}$, find the probability of C using the sample space A. You have reduced the sample space from the original sample space $\{1, 2, 3, 4, 5, 6\}$ to $\{1, 3, 5\}$. So, $\mathrm{P(C|A)} = \frac{2}{3}$

Example:
Let event G = taking a math class. Let event H = taking a science class. Then, $\mathrm{G\ AND\ H}$ = taking a math class and a science class. Suppose $\mathrm{P(G)} = 0.6$, $\mathrm{P(H)} = 0.5$, and $\mathrm{P(G\ AND\ H)} = 0.3$. Are G and H independent?
If G and H are independent, then you must show **ONE** of the following:

- $\mathrm{P(G|H)} = \mathrm{P(G)}$
- $\mathrm{P(H|G)} = \mathrm{P(H)}$
- $\mathrm{P(G\ AND\ H)} = \mathrm{P(G)} \cdot \mathrm{P(H)}$

Note:The choice you make depends on the information you have. You could choose any of the methods here because you have the necessary information.

Exercise:

Problem: Show that $\mathrm{P(G|H) = P(G)}$.

Solution:

$$\mathrm{P(G|H)} = \frac{\mathrm{P(G\ AND\ H)}}{\mathrm{P(H)}} = \frac{0.3}{0.5} = 0.6 = \mathrm{P(G)}$$

Exercise:

Problem: Show $\mathrm{P(G\ AND\ H) = P(G) \cdot P(H)}$.

Solution:

$$\mathrm{P(G) \cdot P(H)}\ =\ 0.6 \cdot 0.5\ =\ 0.3\ =\ \mathrm{P(G\ AND\ H)}$$

Since G and H are independent, then, knowing that a person is taking a science class does not change the chance that he/she is taking math. If the two events had not been independent (that is, they are dependent) then knowing that a person is taking a science class would change the chance he/she is taking math. For practice, show that $\mathrm{P(H|G) = P(H)}$ to show that G and H are independent events.

Example:

In a box there are 3 red cards and 5 blue cards. The red cards are marked with the numbers 1, 2, and 3, and the blue cards are marked with the numbers 1, 2, 3, 4, and 5. The cards are well-shuffled. You reach into the box (you cannot see into it) and draw one card.

Let R = red card is drawn, B = blue card is drawn, E = even-numbered card is drawn.

The sample space $S\ =\ \mathrm{R1, R2, R3, B1, B2, B3, B4, B5}$. S has 8 outcomes.

- $\mathrm{P(R)} = \frac{3}{8}$. $\mathrm{P(B)} = \frac{5}{8}$. $\mathrm{P(R\ AND\ B)} = 0$. (You cannot draw one card that is both red and blue.)

- $\mathrm{P(E)} = \frac{3}{8}$. (There are 3 even-numbered cards, $\mathrm{R2}$, $\mathrm{B2}$, and $\mathrm{B4}$.)
- $\mathrm{P(E|B)} = \frac{2}{5}$. (There are 5 blue cards: $\mathrm{B1}$, $\mathrm{B2}$, $\mathrm{B3}$, $\mathrm{B4}$, and $\mathrm{B5}$. Out of the blue cards, there are 2 even cards: $\mathrm{B2}$ and $\mathrm{B4}$.)
- $\mathrm{P(B|E)} = \frac{2}{3}$. (There are 3 even-numbered cards: $\mathrm{R2}$, $\mathrm{B2}$, and $\mathrm{B4}$. Out of the even-numbered cards, 2 are blue: $\mathrm{B2}$ and $\mathrm{B4}$.)
- The events R and B are mutually exclusive because $\mathrm{P(R\ AND\ B)} = 0$.
- Let G = card with a number greater than 3. $G = \{\mathrm{B4}, \mathrm{B5}\}$. $\mathrm{P(G)} = \frac{2}{8}$. Let H = blue card numbered between 1 and 4, inclusive. $H = \{\mathrm{B1}, \mathrm{B2}, \mathrm{B3}, \mathrm{B4}\}$. $\mathrm{P(G|H)} = \frac{1}{4}$. (The only card in H that has a number greater than 3 is B4.) Since $\frac{2}{8} = \frac{1}{4}$, $\mathrm{P(G)} = \mathrm{P(G|H)}$ which means that G and H are independent.

Example:

In a particular college class, 60% of the students are female. 50 % of all students in the class have long hair. 45% of the students are female and have long hair. Of the female students, 75% have long hair. Let F be the event that the student is female. Let L be the event that the student has long hair. One student is picked randomly. Are the events of being female and having long hair independent?

- The following probabilities are given in this example:
- $\mathrm{P(F\ )} = 0.60$; $\mathrm{P(L\ )} = 0.50$
- $\mathrm{P(F\ AND\ L)} = 0.45$
- $\mathrm{P(L|F)} = 0.75$

Note:The choice you make depends on the information you have. You could use the first or last condition on the list for this example. You do not know P(F|L) yet, so you can not use the second condition.

Solution 1

Check whether P(F and L) = P(F)P(L): We are given that P(F and L) = 0.45 ; but P(F)P(L) = (0.60)(0.50)= 0.30 The events of being female and having long hair are not independent because P(F and L) does not equal P(F)P(L).

Solution 2

check whether P(L|F) equals P(L): We are given that P(L|F) = 0.75 but P(L) = 0.50; they are not equal. The events of being female and having long hair are not independent.

Interpretation of Results

The events of being female and having long hair are not independent; knowing that a student is female changes the probability that a student has long hair.

**Example 5 contributed by Roberta Bloom

Glossary

Independent Events

The occurrence of one event has no effect on the probability of the occurrence of any other event. Events A and B are independent if one of the following is true: (1). $P(A|B) = P(A)$; (2) $P(B|A) = P(B)$; (3) $P(A \text{ and } B) = P(A)P(B)$.

Mutually Exclusive

An observation cannot fall into more than one class (category). Being in more than one category prevents being in a mutually exclusive category.

Probability Topics: Contingency Tables

This module introduces the contingency table as a way of determining conditional probabilities.

A **contingency table** provides a way of portraying data that can facilitate calculating probabilities. The table helps in determining conditional probabilities quite easily. The table displays sample values in relation to two different variables that may be dependent or contingent on one another. Later on, we will use contingency tables again, but in another manner. Contingincy tables provide a way of portraying data that can facilitate calculating probabilities.

Example:

Suppose a study of speeding violations and drivers who use car phones produced the following fictional data:

	Speeding violation in the last year	**No speeding violation in the last year**	**Total**
Car phone user	25	280	305
Not a car phone user	45	405	450
Total	70	685	755

The total number of people in the sample is 755. The row totals are 305 and 450. The column totals are 70 and 685. Notice that $305 + 450 = 755$ and $70 + 685 = 755$. Calculate the following probabilities using the table

Exercise:

Problem: $\mathrm{P}(\text{person is a car phone user}) =$

Solution:

$$\frac{\text{number of car phone users}}{\text{total number in study}} = \frac{305}{755}$$

Exercise:

Problem: $\text{P(person had no violation in the last year)} =$

Solution:

$\frac{\text{number that had no violation}}{\text{total number in study}} = \frac{685}{755}$

Exercise:

Problem:

$\text{P(person had no violation in the last year AND was a car phone user)} =$

Solution:

$\frac{280}{755}$

Exercise:

Problem:

$\text{P(person is a car phone user OR person had no violation in the last year)} =$

Solution:

$\left(\frac{305}{755} + \frac{685}{755}\right)$-$\frac{280}{755} = \frac{710}{755}$

Exercise:

Problem:

$\text{P(person is a car phone user GIVEN person had a violation in the last year)} =$

Solution:

$\frac{25}{70}$ (The sample space is reduced to the number of persons who had a violation.)

Exercise:

Problem:

$\text{P(person had no violation last year GIVEN person was not a car phone user)} =$

Solution:

$\frac{405}{450}$ (The sample space is reduced to the number of persons who were not car phone users.)

Example:
The following table shows a random sample of 100 hikers and the areas of hiking preferred:

Sex	The Coastline	Near Lakes and Streams	On Mountain Peaks	Total
Female	18	16	___	45
Male	___	___	14	55
Total	___	41	___	___

Hiking Area Preference

Exercise:

Problem: Complete the table.

Solution:

Sex	The Coastline	Near Lakes and Streams	On Mountain Peaks	Total
Female	18	16	**11**	45
Male	**16**	**25**	14	55
Total	**34**	41	**25**	**100**

Hiking Area Preference

Exercise:

Problem:

Are the events "being female" and "preferring the coastline" independent events?

Let F = being female and let C = preferring the coastline.

- **a**P(F AND C) =
- **b**$\text{P(F)} \cdot \text{P(C)}$ =

Are these two numbers the same? If they are, then F and C are independent. If they are not, then F and C are not independent.

Solution:

- **a** $\text{P(F AND C)} = \frac{18}{100} = 0.18$
- **b** $\text{P(F)} \cdot \text{P(C)} = \frac{45}{100} \cdot \frac{34}{100} = 0.45 \cdot 0.34 = 0.153$

$\text{P(F AND C)} \neq \text{P(F)} \cdot \text{P(C)}$, so the events F and C are not independent.

Exercise:

Problem:

Find the probability that a person is male given that the person prefers hiking near lakes and streams. Let M = being male and let L = prefers hiking near lakes and streams.

- **a**What word tells you this is a conditional?
- **b**Fill in the blanks and calculate the probability: $\text{P}(\_\_\_|\_\_\_) = \_\_\_$.
- **c**Is the sample space for this problem all 100 hikers? If not, what is it?

Solution:

- **a**The word 'given' tells you that this is a conditional.
- **b**$\text{P(M|L)} = \frac{25}{41}$
- **c**No, the sample space for this problem is 41.

Exercise:

Problem:

Find the probability that a person is female or prefers hiking on mountain peaks. Let F = being female and let P = prefers mountain peaks.

- **a**$\text{P(F)} =$
- **b**$\text{P(P)} =$
- **c**$\text{P(F AND P)} =$
- **d**Therefore, $\text{P(F OR P)} =$

Solution:

- **a**$\text{P(F)} = \frac{45}{100}$
- **b**$\text{P(P)} = \frac{25}{100}$
- **c**$\text{P(F AND P)} = \frac{11}{100}$
- **d**$\text{P(F OR P)} = \frac{45}{100} + \frac{25}{100} - \frac{11}{100} = \frac{59}{100}$

Example:

Muddy Mouse lives in a cage with 3 doors. If Muddy goes out the first door, the probability that he gets caught by Alissa the cat is $\frac{1}{5}$ and the probability he is not caught is $\frac{4}{5}$. If he goes out the second door, the probability he gets caught by Alissa is $\frac{1}{4}$ and the probability he is not caught is $\frac{3}{4}$. The probability that Alissa catches Muddy coming out of the third door is $\frac{1}{2}$ and the probability she does not catch Muddy is $\frac{1}{2}$. It is equally likely that Muddy will choose any of the three doors so the probability of choosing each door is $\frac{1}{3}$.

Caught or Not	**Door One**	**Door Two**	**Door Three**	**Total**
Caught	$\frac{1}{15}$	$\frac{1}{12}$	$\frac{1}{6}$	____
Not Caught	$\frac{4}{15}$	$\frac{3}{12}$	$\frac{1}{6}$	____

Caught or Not	Door One	Door Two	Door Three	Total
Total	____	____	____	1

Door Choice

- The first entry $\frac{1}{15} = \left(\frac{1}{5}\right)\left(\frac{1}{3}\right)$ is P(Door One AND Caught).
- The entry $\frac{4}{15} = \left(\frac{4}{5}\right)\left(\frac{1}{3}\right)$ is P(Door One AND Not Caught).

Verify the remaining entries.

Exercise:

Problem:

Complete the probability contingency table. Calculate the entries for the totals. Verify that the lower-right corner entry is 1.

Solution:

Caught or Not	Door One	Door Two	Door Three	Total
Caught	$\frac{1}{15}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{19}{60}$
Not Caught	$\frac{4}{15}$	$\frac{3}{12}$	$\frac{1}{6}$	$\frac{41}{60}$
Total	$\frac{5}{15}$	$\frac{4}{12}$	$\frac{2}{6}$	1

Door Choice

Exercise:

Problem: What is the probability that Alissa does not catch Muddy?

Solution:

$\frac{41}{60}$

Exercise:

Problem:

What is the probability that Muddy chooses Door One **OR** Door Two given that Muddy is caught by Alissa?

Solution:

$\frac{9}{19}$

Note: You could also do this problem by using a probability tree. See the Tree Diagrams (Optional) section of this chapter for examples.

Glossary

Contingency Table

The method of displaying a frequency distribution as a table with rows and columns to show how two variables may be dependent (contingent) upon each other. The table provides an easy way to calculate conditional probabilities.

Probability Topics: Summary of Formulas

This module provides a review of the probability formulas, including the definitions of independent, complementary, and mutually exclusive events as well as the addition and multiplication rules.

Formula

Compliment

If A and are complements then

Formula

Mutually Exclusive

If A and B are mutually exclusive then ; so

.

If A and B are NOT mutually exclusive then

Formula

Independence

If A and B are independent then:

-
-
-

Probability Topics: Practice with Contingency Tables
This module provides the opportunity for students to apply what they've learned about probability to solve a series of problems given a set of data. Students will practice constructing and interpreting contingency tables.

Student Learning Outcomes

- The student will construct and interpret contingency tables.

Given

An article in the *New England Journal of Medicine* , reported about a study of smokers in California and Hawaii. In one part of the report, the self-reported ethnicity and smoking levels per day were given. Of the people smoking at most 10 cigarettes per day, there were 9886 African Americans, 2745 Native Hawaiians, 12,831 Latinos, 8378 Japanese Americans, and 7650 Whites. Of the people smoking 11-20 cigarettes per day, there were 6514 African Americans, 3062 Native Hawaiians, 4932 Latinos, 10,680 Japanese Americans, and 9877 Whites. Of the people smoking 21-30 cigarettes per day, there were 1671 African Americans, 1419 Native Hawaiians, 1406 Latinos, 4715 Japanese Americans, and 6062 Whites. Of the people smoking at least 31 cigarettes per day, there were 759 African Americans, 788 Native Hawaiians, 800 Latinos, 2305 Japanese Americans, and 3970 Whites. (*(Source: http://www.nejm.org/doi/full/10.1056/NEJMoa033250)*)

Complete the Table

Complete the table below using the data provided.

Smoking Level	**African American**	**Native Hawaiian**	**Latino**	**Japanese Americans**	**White**	**TOTALS**
1-10						
11-20						
21-30						
31+						
TOTALS						

Smoking Levels by Ethnicity

Analyze the Data

Suppose that one person from the study is randomly selected.

Exercise:

Problem: Find the probability that person smoked 11-20 cigarettes per day.

Solution:

$$\frac{35,065}{100,450}$$

Exercise:

Problem: Find the probability that person was Latino.

Solution:

$$\frac{19,969}{100,450}$$

Discussion Questions

Exercise:

Problem:

In words, explain what it means to pick one person from the study and that person is “Japanese American **AND** smokes 21-30 cigarettes per day.” Also, find the probability.

Solution:

$$\frac{4,715}{100,450}$$

Exercise:

Problem:

In words, explain what it means to pick one person from the study and that person is “Japanese American **OR** smokes 21-30 cigarettes per day.” Also, find the probability.

Solution:

$$\frac{36,636}{100,450}$$

Exercise:

Problem:

In words, explain what it means to pick one person from the study and that person is “Japanese American **GIVEN** that person smokes 21-30 cigarettes per day.” Also, find the probability.

Solution:

$$\frac{4715}{15,273}$$

Exercise:

Problem: Prove that smoking level/day and ethnicity are dependent events.

Probability Topics: Homework

This module provides a number of homework exercises related to Probability.

Exercise:

Problem:

Suppose that you have 8 cards. 5 are green and 3 are yellow. The 5 green cards are numbered 1, 2, 3, 4, and 5. The 3 yellow cards are numbered 1, 2, and 3. The cards are well shuffled. You randomly draw one card.

- G = card drawn is green
- E = card drawn is even-numbered

- **a**List the sample space.
- **b**$\mathrm{P(G)} =$
- **c**$\mathrm{P(G|E)} =$
- **d**$\mathrm{P(G\ AND\ E)} =$
- **e**$\mathrm{P(G\ OR\ E)} =$
- **f**Are G and E mutually exclusive? Justify your answer numerically.

Solution:

- **a**$\{\mathrm{G1,\ G2,\ G3,\ G4,\ G5,\ Y1,\ Y2,\ Y3}\}$
- **b**$\frac{5}{8}$
- **c**$\frac{2}{3}$
- **d** $\frac{2}{8}$
- **e** $\frac{6}{8}$
- **f**No

Exercise:

Problem:

Refer to the previous problem. Suppose that this time you randomly draw two cards, one at a time, and **with replacement**.

- $G_1 =$ first card is green
- $G_2 =$ second card is green

- **a**Draw a tree diagram of the situation.
- **b** $P(G_1 \text{ AND } G_2) =$
- **c** $P(\text{at least one green}) =$
- **d** $P(G_2 \mid G_1) =$
- **e**Are G_2 and G_1 independent events? Explain why or why not.

Exercise:

Problem:

Refer to the previous problems. Suppose that this time you randomly draw two cards, one at a time, and **without replacement**.

- G_1= first card is green
- G_2= second card is green

- **a**Draw a tree diagram of the situation.
- **b**>$\mathrm{P}(G_1 \text{ AND } G_2) =$
- **c**$\mathrm{P}(\text{at least one green}) =$
- **d**$\mathrm{P}(G_2|G_1) =$
- **e**Are G_2 and G_1 independent events? Explain why or why not.

Solution:

- **b** $(\frac{5}{8})(\frac{4}{7})$
- **c** $(\frac{5}{8})(\frac{3}{7}) + (\frac{3}{8})(\frac{5}{7}) + (\frac{5}{8})(\frac{4}{7})$
- **d** $\frac{4}{7}$
- **e**No

Exercise:

Problem: Roll two fair dice. Each die has 6 faces.

- **a**List the sample space.
- **b**Let A be the event that either a 3 or 4 is rolled first, followed by an even number. Find $\mathrm{P(A)}$.
- **c**Let B be the event that the sum of the two rolls is at most 7. Find $\mathrm{P(B)}$.
- **d**In words, explain what "$\mathrm{P(A|B)}$" represents. Find $\mathrm{P(A|B)}$.
- **e**Are A and B mutually exclusive events? Explain your answer in 1 - 3 complete sentences, including numerical justification.
- **f**Are A and B independent events? Explain your answer in 1 - 3 complete sentences, including numerical justification.

Exercise:

Problem:

A special deck of cards has 10 cards. Four are green, three are blue, and three are red. When a card is picked, the color of it is recorded. An experiment consists of first picking a card and then tossing a coin.

- **a**List the sample space.
- **b**Let A be the event that a blue card is picked first, followed by landing a head on the coin toss. Find $\mathrm{P(A)}$.

- **c**Let B be the event that a red or green is picked, followed by landing a head on the coin toss. Are the events A and B mutually exclusive? Explain your answer in 1 - 3 complete sentences, including numerical justification.
- **d**Let C be the event that a red or blue is picked, followed by landing a head on the coin toss. Are the events A and C mutually exclusive? Explain your answer in 1 - 3 complete sentences, including numerical justification.

Solution:

- **a** $\{\mathrm{GH,GT,BH,BT,RH,RT}\}$
- **b** $\frac{3}{20}$
- **c**Yes
- **d**No

Exercise:

Problem: An experiment consists of first rolling a die and then tossing a coin:

- **a**List the sample space.
- **b**Let A be the event that either a 3 or 4 is rolled first, followed by landing a head on the coin toss. Find $\mathrm{P(A)}$.
- **c**Let B be the event that a number less than 2 is rolled, followed by landing a head on the coin toss. Are the events A and B mutually exclusive? Explain your answer in 1 - 3 complete sentences, including numerical justification.

Exercise:

Problem:

An experiment consists of tossing a nickel, a dime and a quarter. Of interest is the side the coin lands on.

- **a**List the sample space.
- **b**Let A be the event that there are at least two tails. Find $\mathrm{P(A)}$.
- **c**Let B be the event that the first and second tosses land on heads. Are the events A and B mutually exclusive? Explain your answer in 1 - 3 complete sentences, including justification.

Solution:

- **a** $\{\mathrm{(HHH),(HHT),(HTH),(HTT),(THH),(THT),(TTH),(TTT)}\}$
- **b** $\frac{4}{8}$
- **c**Yes

Exercise:

Problem: Consider the following scenario:

- Let $\mathrm{P(C)} = 0.4$
- Let $\mathrm{P(D)} = 0.5$
- Let $\mathrm{P(C|D)} = 0.6$

- **a**Find $\mathrm{P(C\ AND\ D)}$.
- **b**Are C and D mutually exclusive? Why or why not?
- **c**Are C and D independent events? Why or why not?
- **d**Find $\mathrm{P(C\ AND\ D)}$.
- **e**Find $\mathrm{P(D|C)}$.

Exercise:

Problem: E and F mutually exclusive events. $P(E) = 0.4$; $P(F) = 0.5$. Find $P(E \mid F)$.

Solution:

0

Exercise:

Problem: J and K are independent events. $\mathrm{P(J \mid K)} = 0.3$. Find $P(J)$.

Exercise:

Problem: U and V are mutually exclusive events. $P(U) = 0.26$; $P(V) = 0.37$. Find:

- **a**$\mathrm{P(U\ AND\ V)}$ =
- **b** $\mathrm{P(U \mid V)}$ =
- **c**$\mathrm{P(U\ OR\ V)}$ =

Solution:

- **a**0
- **b** 0
- **c**0.63

Exercise:

Problem:

Q and R are independent events. $\mathrm{P(Q)} = 0.4$; $\mathrm{P(Q\ AND\ R)} = 0.1$. Find $\mathrm{P(R)}$.

Exercise:

Problem: Y and Z are independent events.

- **a** Rewrite the basic Addition Rule $\mathrm{P(Y\ OR\ Z) = P(Y) + P(Z) - P(Y\ AND\ Z)}$ using the information that Y and Z are independent events.
- **b** Use the rewritten rule to find $\mathrm{P(Z)}$ if $\mathrm{P(Y\ OR\ Z) = 0.71}$ and $\mathrm{P(Y) = 0.42}$.

Solution:

- **b**0.5

Exercise:

Problem: G and H are mutually exclusive events. $P(G) = 0.5; P(H) = 0.3$

- **a**Explain why the following statement MUST be false: $P(H \mid G) = 0.4$.
- **b**Find: $\mathrm{P(H\ OR\ G)}$.
- **c**Are G and H independent or dependent events? Explain in a complete sentence.

Exercise:

Problem:

The following are real data from Santa Clara County, CA. As of March 31, 2000, there was a total of 3059 documented cases of AIDS in the county. They were grouped into the following categories (*Source: Santa Clara County Public H.D.*):

	Homosexual/Bisexual	**IV Drug User***	**Heterosexual Contact**	**Other**	**Totals**
Female	0	70	136	49	____
Male	2146	463	60	135	____
Totals	____	____	____	____	____

* includes homosexual/bisexual IV drug users

Suppose one of the persons with AIDS in Santa Clara County is randomly selected. Compute the following:

- **a** P(person is female) =
- **b**P(person has a risk factor Heterosexual Contact) =
- **c**P(person is female OR has a risk factor of IV Drug User) =
- **d**P(person is female AND has a risk factor of Homosexual/Bisexual) =
- **e**P(person is male AND has a risk factor of IV Drug User) =
- **f**P(female GIVEN person got the disease from heterosexual contact) =

Solution:

The completed contingency table is as follows:

	Homosexual/Bisexual	**IV Drug User***	**Heterosexual Contact**	**Other**	**Totals**
Female	0	70	136	49	**255**
Male	2146	463	60	135	**2804**
Totals	**2146**	**533**	**196**	**174**	**3059**

* includes homosexual/bisexual IV drug users

- **a** $\frac{255}{3059}$
- **b** $\frac{196}{3059}$
- **c** $\frac{718}{3059}$
- **d**0
- **e** $\frac{463}{3059}$
- **f** $\frac{136}{196}$

Exercise:

Problem:

A previous year, the weights of the members of the **San Francisco 49ers** and the **Dallas Cowboys** were published in the *San Jose Mercury News*. The factual data are compiled into the following table.

Shirt#	≤ 210	211-250	251-290	290≤
1-33	21	5	0	0
34-66	6	18	7	4
66-99	6	12	22	5

For the following, suppose that you randomly select one player from the 49ers or Cowboys.

- **a**Find the probability that his shirt number is from 1 to 33.
- **b**Find the probability that he weighs at most 210 pounds.
- **c**Find the probability that his shirt number is from 1 to 33 AND he weighs at most 210 pounds.
- **d**Find the probability that his shirt number is from 1 to 33 OR he weighs at most 210 pounds.
- **e**Find the probability that his shirt number is from 1 to 33 GIVEN that he weighs at most 210 pounds.
- **f**If having a shirt number from 1 to 33 and weighing at most 210 pounds were independent events, then what should be true about $\mathrm{P(Shirt\#\ 1\text{-}33 \mid \leq 210\ pounds)}$?

Exercise:

Problem:

Approximately 249,000,000 people live in the United States. Of these people, 31,800,000 speak a language other than English at home. Of those who speak another language at home, over 50 percent speak Spanish. (*Source: U.S. Bureau of the Census, 1990 Census*)

Let: E = speak English at home; E’ = speak another language at home; S = speak Spanish at home

Finish each probability statement by matching the correct answer.

Probability Statements	Answers
a. P(E') =	i. 0.8723
b. P(E) =	ii. > 0.50
c. P(S) =	iii. 0.1277

Probability Statements	**Answers**
d. P(S|E') =	iv. > 0.0639

Solution:

- **a**iii
- **b**i
- **c**iv
- **d**ii

The next two questions refer to the following: The percent of licensed U.S. drivers (from a recent year) that are female is 48.60. Of the females, 5.03% are age 19 and under; 81.36% are age 20 - 64; 13.61% are age 65 or over. Of the licensed U.S. male drivers, 5.04% are age 19 and under; 81.43% are age 20 - 64; 13.53% are age 65 or over. (Source: Federal Highway Administration, U.S. Dept. of Transportation)

Try these multiple choice questions.

The next three questions refer to the following table of data obtained from *www.baseball-almanac.com* showing hit information for 4 well known baseball players. Suppose that one hit from the table is randomly selected.

NAME	**Single**	**Double**	**Triple**	**Home Run**	**TOTAL HITS**
Babe Ruth	1517	506	136	714	2873
Jackie Robinson	1054	273	54	137	1518
Ty Cobb	3603	174	295	114	4189
Hank Aaron	2294	624	98	755	3771
TOTAL	8471	1577	583	1720	12351

Exercise:

Problem: Find $\mathrm{P(hit\ was\ made\ by\ Babe\ Ruth)}$.

- **A** $\frac{1518}{2873}$
- **B** $\frac{2873}{12351}$
- **C** $\frac{583}{12351}$
- **D** $\frac{4189}{12351}$

Solution:

B

Exercise:

Problem: Find $\mathrm{P(hit\ was\ made\ by\ Ty\ Cobb\ |\ The\ hit\ was\ a\ Home\ Run)}$

- **A** $\frac{4189}{12351}$
- **B** $\frac{1141}{1720}$
- **C** $\frac{1720}{4189}$
- **D** $\frac{114}{12351}$

Solution:

B

Exercise:

Problem:

Are the hit being made by Hank Aaron and the hit being a double independent events?

- **A** Yes, because $\mathrm{P(hit\ by\ Hank\ Aaron\ |\ hit\ is\ a\ double) = P(hit\ by\ Hank\ Aaron)}$
- **B** No, because $\mathrm{P(hit\ by\ Hank\ Aaron\ |\ hit\ is\ a\ double) \neq P(hit\ is\ a\ double)}$
- **C** No, because $\mathrm{P(hit\ is\ by\ Hank\ Aaron\ |\ hit\ is\ a\ double) \neq P(hit\ by\ Hank\ Aaron)}$
- **D** Yes, because $\mathrm{P(hit\ is\ by\ Hank\ Aaron\ |\ hit\ is\ a\ double) = P(hit\ is\ a\ double)}$

Solution:

C

Probability Topics: Review

This module provides a number of homework/review exercises related to Probability.

The first six exercises refer to the following study: In a survey of 100 stocks on NASDAQ, the average percent increase for the past year was 9% for NASDAQ stocks. Answer the following:

Exercise:

Problem: The “average increase” for all NASDAQ stocks is the:

- **A**Population
- **B**Statistic
- **C**Parameter
- **D**Sample
- **E**Variable

Solution:

- **C** Parameter

Exercise:

Problem: All of the NASDAQ stocks are the:

- **A**Population
- **B**Statistic
- **C**Parameter
- **D**Sample
- **E**Variable

Solution:

- **A** Population

Exercise:

Problem: 9% is the:

- **A**Population
- **B**Statistic
- **C**Parameter
- **D**Sample
- **E**Variable

Solution:

- **B** Statistic

Exercise:

Problem: The 100 NASDAQ stocks in the survey are the:

- **A**Population
- **B**Statistic
- **C**Parameter
- **D**Sample
- **E**Variable

Solution:

- **D** Sample

Exercise:

Problem: The percent increase for one stock in the survey is the:

- **A**Population
- **B**Statistic

- **C**Parameter
- **D**Sample
- **E**Variable

Solution:

- **E** Variable

Exercise:

Problem:

Would the data collected be qualitative, quantitative – discrete, or quantitative – continuous?

Solution:

quantitative - continuous

The next two questions refer to the following study: Thirty people spent two weeks around Mardi Gras in New Orleans. Their two-week weight gain is below. (Note: a loss is shown by a negative weight gain.)

Weight Gain	Frequency
-2	3
-1	5
0	2

Weight Gain	**Frequency**
1	4
4	13
6	2
11	1

Exercise:

Problem: Calculate the following values:

- **a**The average weight gain for the two weeks
- **b**The standard deviation
- **c**The first, second, and third quartiles

Solution:

- **a** 2.27
- **b** 3.04
- **c** -1, 4, 4

Exercise:

Problem: Construct a histogram and a boxplot of the data.